UNITED STATES DEPARTMENT OF AGRICULTURE
AGRICULTURAL RESEARCH ADMINISTRATION
BUREAU OF AGRICULTURAL AND INDUSTRIAL CHEMISTRY

MINUTES OF CONFERENCE ON TASTE PANEL PROCEDURES AND METHODS OF EVALUATION FOR SOYBEAN OIL AND SOYBEAN OIL PRODUCTS

at the

Northern Regional Research Laboratory
Peoria, Illinois

December 9, 1949

After welcoming those in attendance, Dr. R. T. Milner, Director of the Northern Laboratory, outlined the purpose of the conference. For some time, considerable general interest in taste panel procedures for evaluating edible products had been quite evident. During the San Francisco conference of Food Technology last July, this interest culminated in suggestions for a conference on taste panel procedure and methods of evaluation in the field of soybean oil and its products. Inquiries made to industrial groups operating taste panels indicated agreement as to the desirability of holding such a conference.

Dr. Milner explained that minutes of the present meeting would be written up and sent out to participants. Others interested, but unable to be present, would also receive copies. (These minutes consist of conference proceedings and discussions; texts of papers presented are attached.)

Dr. J. C. Cowan, Head of the Oil and Protein Division, then took over as chairman. After making his opening remarks, he introduced Dr. Stewart T. Bauer of the Drackett Company, the first speaker, under the topic, "Taste Panel Procedures and Methods of Evaluation--General."

After Dr. Bauer's talk on the Drackett taste panel, Dr. Cowan offered a possible explanation of Dr. Bauer's observation that Drackett oils generally received better ratings by the Drackett panel than did oils from other sources. The NRRL panel has found that, in general, laboratory refined and deodorized samples rate higher than commercially refined and deodorized samples. It was thought that perhaps the better vacuum obtainable under laboratory conditions might be largely responsible for this difference. It is possible, Dr. Cowan remarked, that the Drackett panel more often compares a commercially refined oil of another firm with a Drackett crude oil that has been refined in the laboratory.

Dr. Bauer added that fresh samples of oil processed for his taste panel gave some degree of positive correlation between Swift AOM values and taste scores. In answer to a question concerning the colors of various oil samples presented to the Drackett taste panel, Dr. Bauer replied that, in his opinion, members of the panel were not unduly influenced by the colors of the samples rated.

Mr. Robert W. Bates, Armour and Company, then discussed the Armour taste panel. He pointed out that normally only hardened oils were evaluated, and the degree of blandness was the only factor on which the scoring is made; no particular flavors are sought or rated. Mechanics of the tasting procedure are very much like those in the NRRL procedure. However, when samples vary appreciably in color, this difference is masked by presenting samples in a series of different colored shot glasses.

The two persons in charge of the Armour taste panel are not entirely satisfied that all features of the procedure being used are scientific, according to Mr. Bates. Plans are being considered to set up actual threshold tests. The panel is now composed of tasters of relatively long experience, with well-established likes and dislikes. Eventually, persons inexperienced in tasting procedures may be used, to get away from biases. It was observed that it is difficult, apparently, to find 10 individuals who are definitely free of taste bias.

Dr. Edward Handschumaker raised the question whether it was necessary to taste an oil to evaluate its quality. He believes odor is the principal factor for evaluating oils. Taste fatigue sets in rapidly thus reducing the value of flavor responses in the evaluation of oils. If an oil sample is warmed to the proper temperature, its aroma is sufficient to judge the quality. In his experience, the less sensitive observers were aided by the tasting operation. Better observers could rate oils more quickly and accurately by odor. Motions of tasting the samples are gone through, to be sure, but evaluations are primarily based on odors.

One person inquired whether all flavors could be detected through odor. Some flavors were thought to be undetectable by checking aroma only.

Mr. S. J. Rini inquired of Dr. Handschumaker how often, under his testing procedure, would samples rate satisfactory on odor but would subsequently be discarded on flavor. Dr. Handschumaker replied that usually no samples had to be discarded on flavor if odor was satisfactory.

In the talk by Dr. Lester D. Chirgwin, Spencer Kellogg and Sons, Inc., the point of taste bias again came into the discussion. Considerable and consistent individual bias has been observed among members of the Spencer Kellogg taste panel. Complicating the evaluation procedure is the fact that coconut oil is processed in the same plant. In addition to the soybean and coconut oils, cottonseed oil is also tasted. There is a definite division among the panel members between those liking and disliking cottonseed oil when compared to soybean oil. Apparently, manner of processing also affects the responses of various panel members. It is thought that taste bias amounts to as much as 0.3 scoring unit.

Dr. Chirgwin stated that the principal interest of his group is in the quality of the fresh oil and its stability toward reversion and rancidity. Both liquid and hardened oils are evaluated. Tasting is done at the processing plant, the panel meeting before lunch. Any extra tasting sessions are held between 3 and 5 o'clock. Sometimes there are three sessions a day when the number of samples requires. Usually four samples are evaluated at a time. No difference has been noted between morning and afternoon tasting.

Poorer agreement has been found in the taste panel when testing bad oils, according to Dr. Chirgwin. Dr. Cowan made the observation that in either the high or low scoring range, there is usually a good agreement among the NRRL panel members, but a wider range of variance occurs in the middle quality range scoring 5 to 7.

Dr. Chirgwin reported that some commercial lots of oil have been repeatedly checked after they had left the plant. Samples were shipped back for flavor evaluation and their scores were compared with laboratory-aged samples. In this way, some lots of oil were followed for as long as three months. Fair correlation of keeping quality with initial taste score has been found. Good oil at the start usually keeps better. There were, however, occasional outstanding exceptions, according to Dr. Chirgwin.

Mr. Warren H. Goss, Pillsbury Mills, Inc., described the organization of the Pillsbury taste panel, and outlined some of the flavor problems of that company.

The taste panel of A. E. Staley Manufacturing Company, reported on by Dr. Hans Wolff, meets from 11 to 11:30 a.m. Many research samples are tasted; some tests are made on control samples. Report of panel scores are made known to the panel members only after the final calculations have been made. Open panel meetings are held from time to time.

Dr. Wolff explained the screening procedure (small panel tastes all samples, only the interesting ones are passed on to the regular panel for full evaluation) as a means to offset to some extent the major difficulty of having far too many samples to evaluate properly. Some danger of passing up some significant samples by such a screening method was admitted.

Dr. H. C. Dutton inquired whether the effect of light was of practical importance as a factor in the reversion problem. Dr. Wolff stated that its effect should certainly be considered. He believed, however, that the different odors and flavors developed in "light-struck" oils were not a matter of degree of the rancidity or reversion process, but an entirely different factor. Mr. H. G. Spannuth agreed that more information on the effects of light was needed. Dr. Chirgwin and Mr. R. L. Terrill also agreed that "light-struck" oil has a different flavor; an oil can be ruined after only a few hours in the light.

Dr. Handschumaker recounted some experiences of his own on effects of light on shortening. Glass containers had been substituted for metal containers in packaging products. Complaints immediately increased; an odor was noted which differed entirely from ordinary rancidity or reversion. Dr. Handschumaker believes ultra-violet radiation should be checked as the causative agent.

Dr. Wolff asked, at the conclusion of his talk, whether anyone had had experience with freezing oil samples as a means of storage. Mr. Goss reported that his samples are kept in a deep freeze, and that the oils apparently are storing satisfactorily. Miss Helen Moser said that some NRRL samples have been stored for 14 months at 0° F. with no noticeable changes taking place.

Mr. M. M. Durkee, A. E. Staley Manufacturing Company, discussed the flavor problem of soybean oil in the earlier days before soybeans had become a major domestic crop. Mr. Rini inquired as to the progress that had been made in the flavor stability of soybean

oil since the very early days. Mr. Durkee felt that progress had been disappointing. In spite of efforts at better handling, the soybean oil resulting from present-day processing apparently goes through the same undesirable changes as those oils produced by much less careful methods. Antioxidants appear to help considerably in the soybean flavor problem, but in Mr. Durkee's opinion, one is dealing more with breakdown products. Dr. Daubert's work at Pittsburgh was cited as a good illustration of this.

Mr. Rini asked if the present flavor standard of commercial soybean oil is better than that of 10 years ago. Mr. Durkee replied that it was, but not 10 years better!

At the conclusion of a talk on training and control of a taste panel by Miss Moser, NRRL, Mr. Durkee inquired whether there were adverse personal feelings or reactions when persons were discarded from the taste panel because of below-normal performance. Miss Moser answered that no ill feeling had been noted. Dr. Handschumaker remarked that some resentment had been noted in several instances when members had been removed from his panel. He pointed out that in such cases, samples could still be given, but those individuals' responses could be omitted from analysis of data.

Just before the noon recess, a number of the conferees participated in a taste panel conducted in the taste panel room and in accordance with the usual NRRL procedure. Two commercial soybean oils of salad grade and one laboratory-deodorized oil were presented for evaluation. Sample A had a high iron content of 1.50 p.p.m. Sample B had a fairly low iron content of 0.20 p.p.m., while sample C had only 0.015 p.p.m. of iron. All samples were fresh oils that had been stored in the frozen state (0°F) since they were received at the Laboratory. The peroxide values (millimoles oxygen) at the time of tasting were for sample A 0.87; B, 0.56; C, 0.40. On December 5, 1949, the NRRL panel, composed of 13 members, had evaluated these oils in a routine test. Previous to tasting, information was not given the panel that these oils would later be evaluated at the taste panel conference.

Results of the evaluations made by both panels are summarized as follows:

	Sample A (High Iron)	*Sample B (Low Iron)*	*Sample C (Very Low Iron)*
Conferees' Panel:			
Average scores	4.1	7.5	8.2
Standard deviation	1.61	1.0	1.22
Standard error#	.40	.25	.30
Significant differences between samples 1 & 2 **; 1 & 3 **; 2 & 3 +			
Number of tasters (N) = 18			
NRRL Panel:			
Average scores	5.2	8.1	8.6
Standard deviation	1.55	1.05	.89
Standard error#	.46	.32	.27
Significant differences between samples 1 & 2 **; 1 & 3 **; 2 & 3 +			
Number of tasters (N) = 13			

Significant differences between Conferees' and NRRL panels:

Sample A	+	("t" value 1.843)
Sample B	+	("t" value 1.560)
Sample C	+	("t" value 0.967)

\# Used N-2 in formula because of small number of tasters.
\+ No significant difference.
** Results highly significant (1 percent level).

Those participating in the conferees' taste panel were: Philip W. Bateman, Staley; Robert W. Bates, Armour; Stewart T. Bauer, Drackett; Howard C. Black, Swift; Lester D. Chirgwin, Spencer Kellogg; Ronald T. Clause, Procter and Gamble; Frank G. Dollear, SRRL; Maurice M. Durkee, Staley; M. A. Ewan, Clinton; Warren H. Goss, Pillsbury; Edward Handschumaker, Lever Brothers; Joe H. Kirby, Humko; Virgil C. Mehlenbacker, Swift; S. Jack Rini, Kraft; Paul Seaberg, Staley; Hiram T. Spannuth, Wilson; Robert L. Terrill, Spencer Kellogg; and Clarence K. Wiesman, Armour.

Results of the taste panel were discussed briefly at the opening of the afternoon session. Dr. Evans, NRRL, followed with talks on design of experiments and significance of taste tests for small taste panels. In the discussion which followed, Dr. Dutton called attention to the needs for means of identifying different flavors encountered in taste panel work. Qualifying descriptions for tastes should be expanded and standardized. Dr. Handschumaker spoke of a characteristic flavor which develops immediately when interesterification techniques are used in preparing taste samples. This creates a definite bias, and affects directly the value of flavor responses. Some means of controlling the reaction was desirable which would eliminate this flavor development. Dr. Dutton stated that interesterified materials have been prepared at NRRL which gave no characteristic flavor, at least as far as the taste panel had noted.

Methods of evaluating salad oils and mayonnaise were discussed by Mr. Rini, Kraft Foods Company. At the conclusion of his talk, he questioned the validity of direct relationship of threshold values of bitter, sweet, sour and salty flavors with abilities to taste oils; selection of taste panel members based on their taste sensitivities to these flavors may not be on a reliable premise. Some thought, therefore, is being given at Kraft's to ranking tests for selection of oil tasters. Samples of fresh oil would be diluted with graduated amounts of reverted or rancid oil, and the panel members would be asked to arrange the graduated series in proper order.

Mr. Rini inquired whether anyone present had used threshold values of the four "basic" flavors to evaluate abilities of taste panel members. He was informed that such tests had been made with the NRRL panel. Mr. Rini then asked whether any evidence had been obtained which would confirm a relationship between threshold values with the ability to taste oils. Miss Moser recalled that certain tasters who had been discarded on the basis of poor threshold tests during the formation of the first NRRL taste panel had later proven to be some of the best oil tasters.

Dr. Handschumaker attributed this discrepancy to the use of flavor responses only as the basis for selection of panel members. If threshold tests had been made with aromatic compounds, some correlation could have been expected, but because materials such as salt and sugar were used, the lack of odor factors would prevent good correlation. Dr. Handschumaker added that there is a definite need for high caliber individuals on a taste panel to be able to evaluate a product in terms of several attributes simultaneously.

Dr. Wolff inquired whether off flavors in oils show up more markedly in mayonnaise. Mr. Rini replied that he had heard arguments on both sides of the question, but he did not know the answer.

Mr. Durkee inquired what means were used at Kraft in evaluating large shipments of oils. Mr. Rini stated that experienced tasters judged the quality of the oils by objective tests.

Mr. Rini was asked about the average shelf life of mayonnaise. He answered that 4 to 6 months was considered the average range. Mr. Durkee pointed out that the first brand of mayonnaise on the market had had a shelf life of only about 2 weeks. This brand had a reputation for high quality, and to maintain this reputation, it had been necessary for the producer to continually collect the older merchandise because of its inferior keeping characteristics.

Later the tendency was to always blame the vegetable oil for the off flavor in mayonnaise. It was found, however, that eggs were often the cause of undesirable flavors and once the mustard used was found to be the cause. Mr. Terrill mentioned that oil processors get blamed for poor keeping quality of some vegetable oils, but they have had no defense because the real answer is not known.

Dr. Handschumaker pointed out that margarine is also having a difficult time in regard to the shelf life factor. It is removed from sale after 3 weeks. People do not handle margarine on the same basis as butter. If they did, margarine would measure up better. As long as large companies distribute their products on a nation-wide basis, shelf life of the products is a major problem. To meet distribution requirements of large companies, a product must last from 4 to 8 months with 5 or 6 months as an average.

Mr. Goss mentioned that his company had fixed rules for the pickup of their merchandise that had remained on the shelf too long. This rule differed, of course, for the different areas of the country.

Dr. Ronald T. Clause reported on the evaluation methods used by Procter and Gamble Company.

Mr. Rini asked about the correlation of bleach color with flavor when bleach color was used as the grading basis. Dr. Clause stated that for two-thirds of the oils the correlation is good. For the remaining one-third, the color grade was normally lower than the flavor, so the oil was usually up-graded on the basis of the flavor. The flavor rating on liquid oil aged for 2 days is about equal to a flavor rating on hardened oil aged for 3 days.

Mr. Terrill stated that he is in favor of practical scales for scoring oils. He believes this is a good approach and remarked that discussions of the present meeting concern flavor standards that are much above those of the general consumer; however, such standards are not too high for bakers or industrial users. There should be some agreement among those present as to what the flavor standards should be.

Dr. Handschumaker asked members of the group just what they thought the consumer notices when using a vegetable oil product. Mr. Durkee recalled that during the war, seemingly nothing could go wrong with the oil. Every shipment was satisfactory to the customer.

Dr. Cowan asked if the effect of flavor on the individual consumer's acceptance is considered important. Mr. Terrill replied that because the products will wind up in the consumer's hands, the effect is definitely important. Apparently the shortening now on the market is satisfactory to the individual consumer, but is not satisfactory to the industrial users. Dr. Clause agreed that the individual consumer's reaction is important.

Dr. Clarence K. Wiesman stated that this question of consumer standards must be answered. He would like to know how the reactions of a research taste panel would compare with those of the consumer. One additional factor must be taken into account when considering the individual consumer, that is his manner of handling the product. Shortening, for instance, is often held in the home for a considerable time, and receives as much or more abuse there as it does at any other point in its distribution.

Dr. Cowan asked if a consumer-type panel could detect differences in products. Dr. Handschumaker stated that the producer must look to the consumer to find out what he wants and what standards of flavor he expects.

Mr. Kirby pointed out that no type of shortening producer for general consumers contains 100 percent soybean oil. He believes, therefore, that blends of oils should be evaluated.

Dr. Bauer felt that part of the answer to consumer standards is in educating the public to certain flavors. The producer should let the public know when products with flavors different from those to which they are accustomed are to be expected.

Mr. Terrill remarked that a consumer's type of shortening is very likely not being made by the producer. What constitutes good or bad flavor in the opinion of the general consumer? Just where does soybean oil fit in? Are the producers too critical or not critical enough? Does a 7.5 taste score indicate that an oil is suitable to the consumer?.

Dr. Dutton pointed out that the producer should not expect to use the response of a 10-member taste panel to predict the average standards of the consumer multitude. Dr. Cowan asked if it was thought that response from 25,000 consumers would be a good basis to determine standards for soybean oil products. No specific answer to this question was received, but the discussion indicated that even this number might be small unless good population sampling techniques were very carefully followed.

Dr. Handschumaker then discussed at some length the need of the industry for a better means of knowing what the consumer wants.

Dr. Cowan observed that, according to the general discussion, there is no problem about knowing the standards desired by the industrial consumer, but that there is, however, a big problem where the individual consumer is concerned. Dr. Handschumaker agreed that the standards of industrial consumer were known, but that the problem of individual consumer standards should be solved. With this accomplished, the industrial consumers would have less of a problem in marketing satisfactory products of their own.

The statement was made by one participant that the only complaints his sales department brought back on oils were those based on chemical tests. Complaints due to flavor were never received.

Dr. Handschumaker stated that the proper approach to the consumer standard problem was to work on flavor difficulties from the finished product on back, and not from the raw material on out. Also, standards should be set up that can be met. The question still remains, however, as to just how much of the effort necessary to establish consumer standards would be really worthwhile.

Dr. W. R. Fetzer raised the question as to where specifications on vegetable oils are set up. Dr. Handschumaker stated he disapproved of the manner in which the specifications are currently being set because the real importance of various limits in the specification have not been determined properly. The important factors are what the consumer wants and what does the product do. Specifications should be revised from time to time to keep these factors in balance.

Dr. Handschumaker emphasized the fact that the cost of a survey sufficiently broad to determine consumers' standards would be prohibitive for any one individual group. Any such survey must be supported by the entire industry. Such a survey could be conducted because suitable methods are available.

Dr. Cowan then summarized this point of the discussion by stating that there appeared to be general agreement among those present that a need for market research on standards for soybean oil products exists and some action should be considered.

In the course of Mr. H. T. Spannuth's discussion of the procedure used by Wilson & Company in evaluating shortenings, he asked what other research groups were doing to offset the effect of light on flavors of oils and oil products being stored or handled for organoleptic evaluation. Dr. Clause stated that his group used tin containers to keep out the light. In Dr. Handschumaker's laboratory, beakers of low actinic glass (red) were used for the same purpose.

Mr. Spannuth stated that the real problem confronting those in the vegetable oil field was ascertaining what standard of quality or flavor should be set for the public so that the shelf life could be regulated to an optimum period. It is difficult to know, for example, at just what stage of flavor deterioration in potato chips does the public consider unsatisfactory. Certain individuals differ markedly in flavor acceptance. Because this factor is coming to be recognized as a major problem, Mr. Spannuth's group is going more into performance tests instead of panel tests.

In summarizing significant points brought out during the conference, Dr. Cowan reported that several research groups present had stated a desire for a system of interexchanging oil samples for their respective taste panels to evaluate. This arrangement would provide taste references by which the responses of one panel could be compared to those of another. He suggested that the Northern Laboratory would be glad to act as coordinator in such a scheme, and could send out reference samples as desired. He added that perhaps only liquid oils should be sent out at the start. No chemical analyses would be necessary. A flavor evaluation only would be made, and the test results could be sent back to the Laboratory for tabulation and dissemination to the participants under specified arrangements agreed upon in advance.

Dr. Handschumaker pointed out that those entering into such a sample exchange program should agree on one method of evaluation.

Dr. Cowan asked those who wished to enter into an exchange program to sign a list that was circulated among the group. He suggested that the oils to be tested come from commercial firms. It was thought that 5-gallon quantities would be required of each sample, and that those furnishing such samples should hold a 50-gallon drum of the same oil in reserve in case any further testing would be desired, or one particular sample would be selected as a standard. The Laboratory would be responsible for sending out samples and collecting results. Identity of the firms furnishing the various samples would be kept secret.

Dr. Chirgwin inquired whether a selected standard sample would remain constant in flavor. Someone suggested that the oil standard could be placed in a freezer, using food locker plants near the Laboratory, if necessary.

The question was brought up as to whether evaluations of sample oils should be performed by the participants on the same day. It was decided that if the oils remained frozen while in transit and until tested, this factor would not be a problem.

It was suggested that about eight samples per year would be about right for reference evaluations. Each sample would be graded by the NRRL panel, and by each of the other laboratories, and a comparison made of the flavor scores. A ranking evaluation would be wanted, at least, and perhaps a number score on each of the samples would be desirable. Later on, if an exchange system of liquid oils works out, then a similar procedure on other vegetable oil products could perhaps be used.

Mr. Spannuth raised the question whether other oils in addition to soybean oil should be used in the sample exchange program. Dr. Cowan felt that soybean oil should be the only standard at first, unless there was a demand for a cottonseed oil standard as well. Dr. Handschumaker thought that cottonseed oil should be used as the standard with unknown samples of soybean oil. Three oils, corn, cottonseed, and soybean, might possibly be used.

Mr. Rini pointed out that the sample exchange procedure should be aimed at standardizing panels, not oils.

Dr. Cowan concluded the conference by stating that the Laboratory would begin working on details of a method to circulate standard exchange oils. Two points in the exchange program were generally agreed upon; (1) score sheets similar to those of NRRL would be used, and (2) there would be no need to store standards, at least in the early phases of the exchange program.

I. TASTE PANEL PROCEDURES AND METHODS OF EVALUATION--GENERAL*/

Procedures and Methods Used by The Drackett Company---Stewart T. Bauer, The Drackett Company. ***(Written from notes taken during the Conference)***

The Drackett Company produces soybean oil in the crude form only. It has, therefore, no plant deodorizing, refining, or bleaching equipment. Taste samples are laboratory produced, the refining being done in pilot-scale equipment, and the deodorizing in a glass deodorizer (Bailey type) which, except for some minor changes, is similar to the apparatus used by the Northern Regional Research Laboratory. These minor changes permit the addition of materials during the deodorization process. With this provision, a determination can be made of the effects on the flavor stability of various substances added at different stages of deodorization.

Drackett's taste panel consists of 15 or 16 persons, from all branches of the organization. Its members range from the president of the company on down. The panel usually meets two days in succession during which time four to six samples, in duplicate, can be evaluated. Meetings of the panel are held in the morning, between 9:30 and 11:00. Since the panel is comprised of men from two different laboratories, taste samples are transported from one laboratory to another during a taste run.

Procedure used in the tasting operation follows quite closely that of the Northern Regional Research Laboratory. Volume of an oil sample is 7.5 ml.; samples in beakers are held at 55°C. in wooden blocks. Two samples, which are coded, are rated at a time, first on odor, then on flavor. Tasting is done in the order of the least to the most objectionable odor. Scoring is on the basis of 1 to 10, with the best odors and flavors receiving the highest scores. To designate the scores given, descriptions of odors and flavors are entered in proper spaces of the score sheet instead of check marks. "Rewards" are furnished at irregular intervals.

No elaborate statistical method is used for the evaluation of the responses. Repeat samples are run through the panel from time to time and good checks are obtained. One member of the panel has been taken off, however, because his responses were too different from those of other panel members.

Mr. Bauer stated that, at Drackett's, the taste panel has proved itself to be a very useful analytical tool. One point of concern however, has been the tendency of the panel members to favor Drackett oils in their flavor rating. Mr. Bauer voiced a desire to have some means of standardizing his panel; some means of comparing its responses to control samples evaluated by other panels.

Armour Flavor Panel---Robert W. Bates, Armour and Company:

The flavor panel meets in the New Products Kitchen under ideal physical conditions. The lighting, air conditioning, etc., are most modern. Two graduate Home Economists conduct the panel.

---

*/ Presented at the ***Conference on Taste Panel Procedures and Methods of Evaluation for Soybean Oil and Soybean Oil Products*** held at the Northern Regional Research Laboratory, Peoria, Illinois, on December 9, 1949.

The panel, in general, is patterned after the one used in the Northern Regional Research Laboratory Some modifications, however, seem necessary in an industrial organization.

The panel is made up wholly of interested persons and has been in existence about 2-1/2 years The personnel includes representatives from the Operating, Sales, Laboratory, Bakery, and Quality Control Departments. Every member of the panel is interested in oil flavor.

The panel meets at 11 a.m and 2 p.m. and four samples (one a standard) are tasted at one time. This number seems optimum for panel accuracy and maintaining schedules. After all have tasted the oils the relative scores are inspected and the results discussed. The panel consists of about ten persons.

We found it absolutely necessary to procure colored glass containers for the oils as quite frequently the standard will be lighter in color (or vice versa) and panel members can detect the standard from the color. We use blue, purple, yellow and green liquor glasses (about 3/4 oz. capacity) for this purpose.

The results on the 11 a.m. panel usually show greater uniformity (a smaller standard deviation) than do those on the 2 p.m. panel, and in our opinion oils should be tasted before lunch whenever possible.

The temperature of the oils when tasted is between 50° and 55°C. and a "wash out" portion which must be an extremely bland oil is given each taster. We have found this to be a successful procedure.

Our grading system is as follows: Excellent - 9-10; Good - 7-8; Fair - 5-6; Poor - 3-4; Very Poor - 1-2. We attempt to maintain a standard of 7 minutes on our production.

On reversion studies we incubate the oils at 60°C. in the dark, in stoppered 4 oz. oil bottles and taste them at 0, 3 and 5 days.

A reward system is followed after the panel, which may be coffee, cake, etc. With our panel of interested personnel a reward system is probably not necessary but with a panel composed of persons not vitally involved in oil processing a reward system would be important.

Our greatest disagreement between panel members is on oils in the 6-7 range. There is significantly less disagreement on oils of excellent and poor flavors. We have found smokers to be as effective tasters as non-smokers.

Outline History of Taste Panel for Edible Oils at Edgewater Plant—Lester D Chirgwin, Spencer Kellogg and Sons, Inc.:

I. Initial Panel: 1943-4

A. Purpose: production control

B. Panel members

1. Control chemist tasted each batch
2. Supervisory personnel (Supt. and Asst. Supt.) checked questionable batches; arbitrarily spot checked at convenience
3. Occasional visitors (company sales representatives, main office personnel, central control lab members)
4. Plant operators, on occasion

C. Scoring: on basis of four grades:

4 - excellent	2 - fair
3 - good	1 - poor

D. Recording: no systematic correlation or recording of grades; a batch was either shipped as acceptable or reprocessed

II. First Modification: 1944-5 (Only the scoring was modified, by spreading the numerical range)

A. Scoring: on basis of seven grades

6 - excellent	2 - fair
5 - good to excellent	1 - poor to fair
4 - good	0 - poor
3 - fair to good	

B. Purpose: panel-members, and recording were unchanged except for increased frequency of participation by others than the control chemist

III. Final Modification: 1947 to the present

A. Adopted Northern Regional Research Laboratory scoring system (but with practically none of the physical improvements; i.e., elimination of color factor, air conditioning, etc.)

B. Applied system to product control and an intensive program for improvement of flavor quality of soya margarine oils

C. Panel Members

1. For control work; same as under I
2. For margarine improvement program, a standard 4-member panel (works manager, supt. and asst. supt. of edible oils dept., supt. devel. lab.) augmented arbitrarily by visitors and other employees, as before

D. Recording

1. For control work, original score sheets were filed with shipping papers
2. For margarine oil improvement program, all individual and averaged grades were tabulated with the sample history and comments

IV. Description of Margarine Oil Improvement Program

A. Frequency of oil grading: 1 to 3 times daily (generally 2)

B. Time: 11 a.m. to 1 p.m. (preferred)
3 p.m. to 5 p m
9 a.m. to 11 a.m.)
2 p.m. to 4 p.m.) used when necessary, but least desirable

C. Samples per series: 1 to 8; 4 predominately

D. Sample presentation

1. About 30 ml. of oil in 50 ml. beakers, covered with watch glasses and coded
2. Held between 125° - 150°F. in a heated aluminum block bored to hold 8 samples
3. The single set of samples was graded by each panel-member, oil was transferred to and tasted from stainless steel spoons

E. Sample Evaluation

1. Preliminary classification by odor if possible: odor was generally recorded (optional) but not used in final evaluation
2. Samples tasted in order of apparently decreasing quality (odor basis)
3. First sample retasted to overcome "shock effect" of first, or "palate-conditioning" taste
4. Each panel member allowed to flavor by his own preferred method (i e. large or small doses, swallowing, rinsing, etc., all optional)

F. Grading

1. After extensive grading of one type of oil, whole-number grades were refined by the use of plus and minus signs, and by grading on the line between two categories
2. Numerical averaging of such grades was accomplished by adding or subtracting 0.3 or 0.5 to the unit values (i.e. 9 + = 9.3; 9- = 8.7; a grade half-way between 9 and 10 would be 9.5)

G. Panel Discussion

1. Immediately after the last panel-member finished grading a series, results were tabulated and the final values discussed along with the sample history
2. Any panel-member who "went off" on a grade would often retaste the sample in question, but was not allowed to change his initial grade

H. Panel Proficiency

1 Proficiency has been evaluated four times since adoption of NRRL scoring method
2. Proficiency was found to improve slightly with experience and to be somewhat higher for regular panel members than occasional graders
3. The following values are an indication of panel proficiency (based on tests on 700 samples):

Total individual grades within one unit of the panel average	- 93%
Ability of individuals to place samples in same order as indicated by panel average (at 0.05 level of significance)	- 95%
Overall precision of individual panel members	- ±0.5 units
Panel standard deviation	- ±0.7 units

4. In general, agreement was poorer on low quality oils and on oils other than hydrogenated soya
5. Some individual bias was apparent in that those responsible for the production of the oil tended to grade cautiously, or on the low side (i.e. were hypercritical)

I. General Notes

1. Odor was found to be completely unreliable in cases of good quality oils. Where odor was at all appreciable, however, quality was generally quite low.
2. Quality of all oils was found to decrease slightly during storage and shipment, but higher quality oils generally displayed better stability than those of low quality.

Taste Panel Procedures at Pillsbury Mills, Inc.— Warren H. Goss, Pillsbury Mills, Inc.:
(Written from notes taken during the Conference)

Pillsbury is interested in the evaluation of fats and oils because it is a large purchaser of a wide variety of such materials for incorporation into premixes. Premixes have two general outlets, the commercial bakeries for which large volume lots of the bulk premix are produced, and the grocery store for which the premix product is packaged in small lots. In the bulk premix field, turnover of stocks is rapid (normally a matter of a few months only) and reversion is not a serious problem. Grocery premix products are a different matter, however, as packages, on an average, stand on the shelf much longer. The standard required of such products is considered to be 12 months of storage at 100°F. When subjected to such a standard, an oil with any tendency towards reversion is not used because off-flavors cannot be tolerated in a product which must be entirely satisfactory to the consumer. Soybean oil is one oil that tends to become both rancid and reverted.

Fats and oils are purchased on specification. It is difficult to write flavor stability into specification, so the taste panel has been established to assist in the purchasing of such materials. It is the desire of the company to have the taste panel conform with others in the field that are based on scientific procedure.

At Pillsbury, an expert on the evaluation of baked goods exemplifies the difficulty of attempting to establish a scientifically-operated taste panel for products other than fats and oils. This particular individual is highly qualified in his field and, by resorting to methods which might be considered unscientific by many, is able to evaluate an enormous quantity and variety of samples in a very short time. Despite the unscientific procedure employed much of the time, he can detect quite consistently the inception of reversion, rancidity, and other forms of deterioration in stored bakery products long before the degradation can be observed by a panel using more scientific methods.

Taste panel organization and procedure at Pillsbury are almost identical to those features of the Northern Regional Research panel. The oil taste panel is quite new, and, as yet, no conclusions can be drawn as to its accuracy and dependability. The oil panel, which is not used for other products, usually consists of about 12 members. Tasting booths are similar to those at Northern Regional Research Laboratory. Air from the booths is recirculated through carbon absorption system and colored lights are used to illuminate the booths. Booths are designed for use in evaluating baked goods as well as oils.

At Pillsbury, there are ample baked goods around to serve as "rewards".

Mr Goss gave advance notice of his desire to have some system initiated for exchange of oil samples between panels for panel standardization and comparison purposes.

Taste Panel Procedures of the A E Staley Manufacturing Company——*Hans Wolff A. E. Staley Manufacturing Company:*

The taste panel for the evaluation of oil flavors was set up at the research laboratory of the Staley Company according to the method described by Moser, Jaeger, Cowan, and Dutton (J. Am Oil Chem. Soc. *24* 291 (1947)). The object of this outline is to describe variations from the taste panel procedures of the Northern Regional Research Laboratory.

Since flavor unstability is usually not detectable for two to three weeks we were interested in accelerated aging of the oils Our first attempt to accomplish this by storing the oil in a laboratory oven at 60°C. proved unsatisfactory After about 24 hours the oils acquired a strange flavor unfamiliar to the tasters, regardless of whether or not air was passed through the dark oven The unusual flavor made evaluation very unreliable.

By using light in an oven rapid taste evaluation was possible. The oven, of dimensions 30"x24"x30", has two rotating discs 9-1/2 inches apart and is lighted with two G.E. 6-watt daylight bulbs, each placed 3-1/2 inches from the edge of each disc. In order to accomplish even exposure to light, the oil samples in 8 or 16-ounce bottles, loosely capped, are placed on the rotating disc. The disc moves at 100 r.p.m. The temperature is kept at 60°C. ± 1. Forced air circulation is provided by a fan; the oven is not air-tight. The following table shows a comparison of the flavor scores of the same starting oils, stored in the oven and at room temperature.

	Stored at 60°		Stored at Room Temp.	
Samples	12 hrs.	24 hrs.	13 days	20 days
A	6.8 ± 0.4	5.2 ± 0.4	6.8 ± 0.4	5.1 ± 0.4
B	5.8 ± 0.4	5.2 ± 0.4	6.0 ± 0.7	5.1 ± 0.4
C	6.3 ± 0.8	5.3 ± 0.4	6.3 ± 0.8	5.0 ± 0.4
D	5.8 ± 0.4	4.9 ± 0.5	6.3 ± 0.4	4.9 ± 0.4

Only five tasters evaluated these oils; our regular taste panel has 17 members, and usually 11 to 13 appear. The small taste panel of five to six members is used for the screening of samples; the final evaluation is carried out by the big panel.

This rapid aging is a useful tool for screening research samples We believe that flavor unstability (reversion) can be detected quite satisfactorily by this method but that it is not reliable for testing development of rancidity. The peroxide values are also difficult to correlate between fast aging and room temperature aging, but little work has been done on this.

Recently we replaced the small beakers used for tasting by 3.4-ounce souffle cups. These cups appear to be quite satisfactory and eliminate the cumbersome washing of large amounts of glass cups. The oils are heated to 40°C., prior to tasting, in glass bottles and poured into the souffle cups immediately before tasting. Whereas most other paper containers examined for this purpose contained objectionable waxes or other materials, these souffle cups appear to be very satisfactory.

Our tasters experienced repeated difficulties in orienting their scoring We have therefore given the taster a control sample to which a flavor score of seven is given Since most of our evaluations concern laboratory experimental oils we use a laboratory refined and deodorized sample for this control The control sample is a freshly deodorized oil and usually prepared from the same crude oil as the experimental samples.

Oil Tasting Problems—M M Durkee A E Staley Manufacturing Company (Written from notes taken during the Conference)

In the past the flavor problem of soybean oil has been very serious Back in 1919 when considerable quantities of Manchurian soybean oil were available refining equipment had not reached its present stage of efficiency Manchurian soybean oil was very prone to develop a fishy flavor and odor Methods were sought to blend this soybean oil with cottonseed to produce a passable edible item. There was, however, no guide as to consumer acceptance of flavor.

Mr Durkee recounted his experience in the olive oil trade where oil flavors cause considerable trouble Olive oils with widely differing flavors were customarily blended to match the flavor of a standard product Variety age of the tree method of processing, locality where grown, and possibly other factors all contributed to the widely differing flavors encountered in the olive oils on the market.

A taste panel of six experts checked olive oil shipments on a fee basis These men were adept at detecting adulteration (usually peanut oil); samples from almost every barrel in a shipment had to be tasted To prepare for a job, the members ate no breakfast and only a cup of hot tea was taken to "clear the taste" As many as 50 to 100 samples at a time were sometimes checked. The acuity of these tasters was such that they could often determine the areas which produced the olives from which the oils were obtained.

In 1928, when U S soybean production got under way, the bad memories of oil produced in 1920-21 delayed interest of the trade in domestic soybean oil An evaluation method was needed. A method of testing quality made use of a 12-ounce bottle in which was placed a small amount of the oil. The bottle was tightly corked and the air over the oil saturated with odor by vigorous shaking of the bottle Smell and taste went together so evaluation of the oil was based on the odor To intensify the smell samples were heated to 425°F , and the odor was checked while still hot This method was satisfactory because after all, only observable differences were being sought.

Oil from domestic soybeans processed better than Manchurian beans Although practical observations only were the basis of determination light was known to be one of the worst enemies of soybean oil No research was done on the subject to clarify the problem

Mr. Durkee believes that foreign oil might be used as one type of a standard for taste panel work. Oil used as a standard should be fresh, fully refined and deodorized. Perhaps standard samples could be prepared from any one or all of a group of oils such as cottonseed, sesame, peanut, or sunflower.

11 TRAINING AND CONTROL OF A TASTE PANEL*/--Helen A Moser, NRRL

Work on the flavor stability of soybean oil was started at this Laboratory in 1944 Realizing that existing physical and chemical tests were not suitable for measuring the off-flavor development in soybean oil, organoleptic methods for evaluation were investigated Many edible products already were being evaluated by means of analytical and consumer taste panels Because of the nature of the research it was necessary to set up an analytical type of taste panel Since the oil would be given various treatments it was important to have that type of panel which could detect small differences between treated and untreated oils For this reason the selection training sensitivity and consistency of the individuals comprising the panel were of extreme importance.

The first taste panel was selected following the procedure of the Bureau of Human Nutrition and Home Economics, United States Department of Agriculture (1) Briefly a group of people were given preliminary acuity tests for the primary tastes salt sweet sour and bitter. After eliminating those who had very high thresholds or made incorrect identifications, the remaining group was given a second series of tests as a check on their performance It was from this group that the first panel was selected.

In 1946 the taste panel was reorganized and up to the present time there have been few changes in procedure. However, the physical conditions under which we operate the taste panel have been improved. Figure 1 shows the floor plan for the entire room This room has a constant temperature of 78°F and a relative humidity of 40 per cent Turning on the pilot light at the booth signals the attendant in the preparation area and she passes the samples through the sliding doors to the panel member who tastes the samples and records his evaluations on the standard score sheet After tasting he leaves through the office where he can see what the samples were that he tasted and compare his evaluations with those of other taste panel members. Here he also receives his "reward" - cookies - which helps to remove any lingering flavors of oils from his mouth This opportunity to discuss the samples lends much to maintaining interest in the work Meetings of the entire taste panel are held often to give them the results of completed work All these things tend to create interest in the work and we feel that this is extremely important in holding a panel together Bengtsson states in a recent paper that "interest is the prime attribute of a good taster" (2).

Back of the taste panel booths is the preparation area Here the samples are stored in the dark in an oven at 60°C for accelerated storage If the samples are to be used immediately they are placed in the refrigerator which is held at -12°C Otherwise samples are stored in a cold room held at 0°F The hot plate for warming the samples is imbedded in the regular laboratory bench The office space is at the right of this area. Thus all taste panel operations from preparation through tasting and statistical analysis are carried out in this one room.

---

*/ Presented at the ***Conference on Taste Panel Procedures and Methods of Evaluation for Soybean Oil and Soybean Oil Products*** held at the Northern Regional Research Laboratory, Peoria, Illinois, on December 9, 1949.

Substance of this talk has been previously presented:
1. J. Am. Oil Chem. Soc., 24(9), 291-296. 1947.
2. Talk to The Institute of Food Technology, July 1949 meeting.

Because some of the experienced taste panel members left the Laboratory and because of absences, it was necessary to have trained substitutes available so that a uniform number of tasters can be maintained and thus assure us of more reliable results. Feeling that perhaps the threshold tests for the primary tastes were not important in oil tasting it was planned to train a new group of individuals by using oil samples The training of this new group was purely experimental as there were no prescribed procedures for oil tasting set up by industry or research organizations. A three weeks' training period was used to familiarize the training panel with procedures and with the flavors of the oils with which they would be working. While it was felt that a longer training period would have been desirable, a good agreement in results was shown.

The volunteer trainees met and the purpose of taste panels was explained. The two types of panels, consumer and analytical, were described with emphasis on our panel as an analytical panel and its work in evaluating samples from varied refining procedures. The tasting procedures were described, the equipment shown, and the score sheet explained and discussed. To become familiar with the taste sensation, odor, and flavor of oils, as well as the tasting procedure, all members of the training panel met daily for round table discussions. All members smelled the samples, and the odors were discussed and recorded on a large score sheet on the wall. The samples were then tasted and the taste sensation and flavors discussed. Scoring of the oils was carried out in this way in an effort to develop similar reactions to the oils. All members seemed to recognize differences in mineral, cottonseed, and soybean oil which were given them. Various samples were presented for practice scoring. Good agreement of the panel in scoring samples presented "blind" was shown by very small standard deviations from the means.

Open discussions were held again in order to become familiar with the flavors and off-flavors in stored soybean oils. Members tasted samples, discussed the predominating flavors and finally agreed on such flavors as buttery, beany, rancid, grassy, painty, etc. Memory associations seem to be very important at this point and so those who were very unfamiliar with oils benefited by the discussion and the suggested flavor descriptions. As an example, one treated oil had a peculiar off-flavor which many could not describe and which many hesitated to describe as they really wanted. They thought it tasted like "watermelon" but felt that such a flavor could not be present in oil. However, when the sample was openly discussed those who had difficulty in identifying the sample readily agreed with those who were brave enough to say that it tasted like watermelon or cucumbers. Again, numerous samples of oils were given for practice scoring and flavor description.

After three weeks of such training, the new group agreed quite well in their description of flavors and in their numerical evaluations of the samples. The group seemed to take great interest in improving their tasting ability and looked forward to their daily "report cards" which gave the panel means and their individual scores for the oils tasted on the previous day.

In order to utilize the services of the new group to the best advantage, it was necessary to measure individual as well as panel performance. By giving the same tests to both the training and the experienced panels, it would then be possible to evaluate the results statistically and on this basis select a taste panel whose precision should improve.

Two methods of analysis were used to measure panel performance and individual performance. The first, the control chart method which measures the reproducibility of the individual's scoring on a single oil, is shown in Figure 2 The same oil was presented 20 times and the results supplied the data from which the control charts were drawn. The discontinuous lines for the limits in these charts result from the allowance made for the unequal number of tastings by panel members From the chart showing averages for the training panel we see only two members were outside the limits while five of the experienced panel members were outside the limits. This indicates that some of the experienced taste panel members were grading the oil samples high and some were grading them equally as low The control chart on standard deviations shows that each member was consistent in his grading of the samples whether he graded high or low but that the panel as a whole was not consistent in its thinking as to the quality of that oil sample. It is interesting to note that taster No. 7 on the control chart of standard deviations for the training panel was outside the upper limit and his scoring was erratic. That same taster was outside the limits on the control chart of averages. The fact that only one member was outside the upper limit on the control chart of standard deviations for the training panel may indicate that the group as a whole benefited by its discussions concerning the evaluation of samples during the training period and, therefore, the members were reacting more along the same lines. Tasters No. 8 and 10 on this same chart have small standard deviations and may be considered very reliable tasters.

The second method for determining individual and panel performance was the correlation and regression method, used to measure the ability of the individual to distinguish differences between oils. Twenty-two paired samples of oil were presented at various times throughout the test period The samples were made up of control samples and pairs of oils which had been stored 1, 3, 7, and 15 days at 60°C. The individual's score was correlated with the average score of the remainder of the panel. The correlation coefficients for the training panel ranged from 0.55 to 0.88 while those of the experienced panel ranged from 0.56 to 0.90. The regression coefficient, which shows how many units the individual score changes for each unit change in the panel average, ranged from 0.62 to 1.84 for the training panel and 0.40 to 1.33 for the experienced panel In Figure 3, scatter diagrams are shown for the tasters from both the training panel and the experienced panel who had the highest and lowest correlation coefficients Such diagrams show at a glance the relationship between the two variables, the individual score and the panel average; they also show the amount of change in individual score for each unit change in the panel average; how nearly the estimated values agreed with the values actually observed for the variable being estimated; and the accuracy with which predictions can be made under the same conditions.

To determine the individual's ability to differentiate between two samples, the "triangular test" was used. In this test three samples were presented to the taster. Two were identical while the third was different Each time this test was presented the panel members were asked to select the two samples which were identical. The degree of difference in these two samples is of great importance If the difference in the oils is too great, it would be detected by even the poorest tasters. If the difference is too slight, chance selection alone could give correct answers once out of every three trials. When the number of tests is small, one cannot assume that chance will produce exactly 33 percent of correct answers. It is necessary in such cases to know how far the number of correct answers must exceed 33 percent before one may be certain

that guessing is eliminated. This can be calculated according to Bengtsson's adaption of the chi-square analysis to the triangular test (3, 4). Table I (see page 5) shows the results of the triangular test presented to the training and the experienced panels. Experience seems to have had some influence on this test as the results of all but two of the experienced tasters are significant while the results of all but five training panel members are significant.

Some interesting observations may be made from the results of these tests. Taster No. 7 on the training panel had the lowest correlation coefficient, was unable to distinguish the difference in samples in the triangular test, was outside the limits on the control chart of averages and was erratic in the scoring of identical samples. Such results furnish sufficient evidence for dropping taster No. 7 from the taste panel group. Another interesting case is that of taster No. 7 on the experienced panel. Although he had the highest correlation coefficient of 0.90, he had one of the lowest percentages of correct answers in the triangular test. This is difficult to explain since this particular individual has a very keen sense of smell and is able to identify many chemical compounds by odor and to identify the various components of perfumes. It was evident throughout, however, that there was a psychological reaction each time this test was given and perhaps the factor of confusion caused such a response. Another explanation is that this same taster was unable to detect diacetyl (used in this triangle test) in low concentrations, and perhaps, therefore, he could not detect the buttery flavor in the oil. The control charts showed that while this same individual graded slightly higher than the panel, as a whole he was consistent in his grading.

After a careful study of the results of these tests, a panel of 18 persons was chosen from the entire group of 22. After a 4-month period of regular tasting the results of this group were evaluated by the correlation-regression method. With the exception of two persons, the panel showed good agreement. In the cases of those two persons, known causes accounted for their slightly-below-average performance. Recently, control charts were drawn from data collected over a period of 14 months. Again, the panel showed good agreement with the exception of two persons; one graded slightly higher than the panel average and one graded slightly lower than the panel average. These two persons will be watched closely for future performance and may be given further training. On the control chart for standard deviations there was but one person who showed erratic scoring.

Table II (see page 6) shows some typical results of one run. The samples as they came from the deodorizer were about the same in quality and were rated good. Ten is the highest possible score and is reserved for bland oils. After these oils were stored at 60°C for four days they were again evaluated by the paired-sample technique. Off-flavors had developed as can be seen from the scores assigned the oils. The consistency of the panel is shown by the panel averages each time the oil was presented paired with one of the remaining samples.

The job classifications of the present taste panel members may be of interest. The panel consists of a bookkeeper, a technical editor, 8 chemists, a technical analyst, 3 chemical engineers, 2 agronomists, and 2 administrators. Type of work apparently has no influence on the taster's ability to judge oils, for the statistical analysis shows that a chemical engineer and an administrator had the highest correlation coefficients. That smoking has not affected the results of the taste panel is indicated by the fact that one of these two tasters is a chain smoker and the other is a non-smoker.

TABLE I

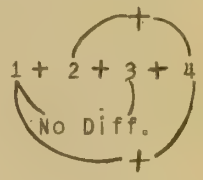
Results of Triangular Tests

Taster	No. Times Tasted	No. Correct Responses	% Correct Responses	Significance of Results 1/
		Training Panel		
1	18	11	61.1	*
2	17	10	58.8	*
3	18	11	61.1	*
4	15	7	46.7	+
5	15	5	33.3	+
6	16	10	62.5	*
7	19	9	47.4	+
8	20	13	65.0	**
9	18	12	66.7	**
10	18	6	33.3	+
11	18	9	50.0	+
		Experienced Panel		
1	20	19	95.0	***
2	18	14	77.8	***
3	16	11	68.8	**
4	18	14	77.8	***
5	19	11	57.9	*
6	14	9	64.3	*
7	19	7	36.8	+
8	15	14	93.3	***
9	17	15	88.2	***
10	18	12	66.7	**
11	18	9	50.0	+

1/ + no significant difference; * slightly significant; ** significant; *** highly significant.

TABLE II

Taste Panel Evaluation

1 Raffinate Plus .02% Phosphatides	2 Raffinate Control	3 Original Oil Plus .02% Phosphatides	4 Original Oil Control	Sig. Dif.
		0 Time		
8.7	8.4	8.7	8.6	
		After 4 days Storage at 60° C.		
6.9	5.8			*
6.6		6.9		+
6.6			6.6	No Diff. in Means
	5.9	7.4		**
	6.2		6.5	+
		7.5	6.8	+

From the work done at this Laboratory, a few general observations might be made. First, that through a well-planned training program, most individuals can be trained to become oil tasters. Second, that interest in the work is very important. Third, that regularity, quiet and comfortable surroundings, orderliness, and smoothness of presentation contribute to more accurate and consistent evaluations. Fourth, that statistical evaluations such as those described enable the director of a taste panel to watch for assignable causes of variation in individual as well as in panel performance. Finally, since there is no known chemical or physical test to detect off-flavor development in soybean oil, it seems clear that the organoleptic method must be used and that any refinement or improvement in methodology is important to this line of research.

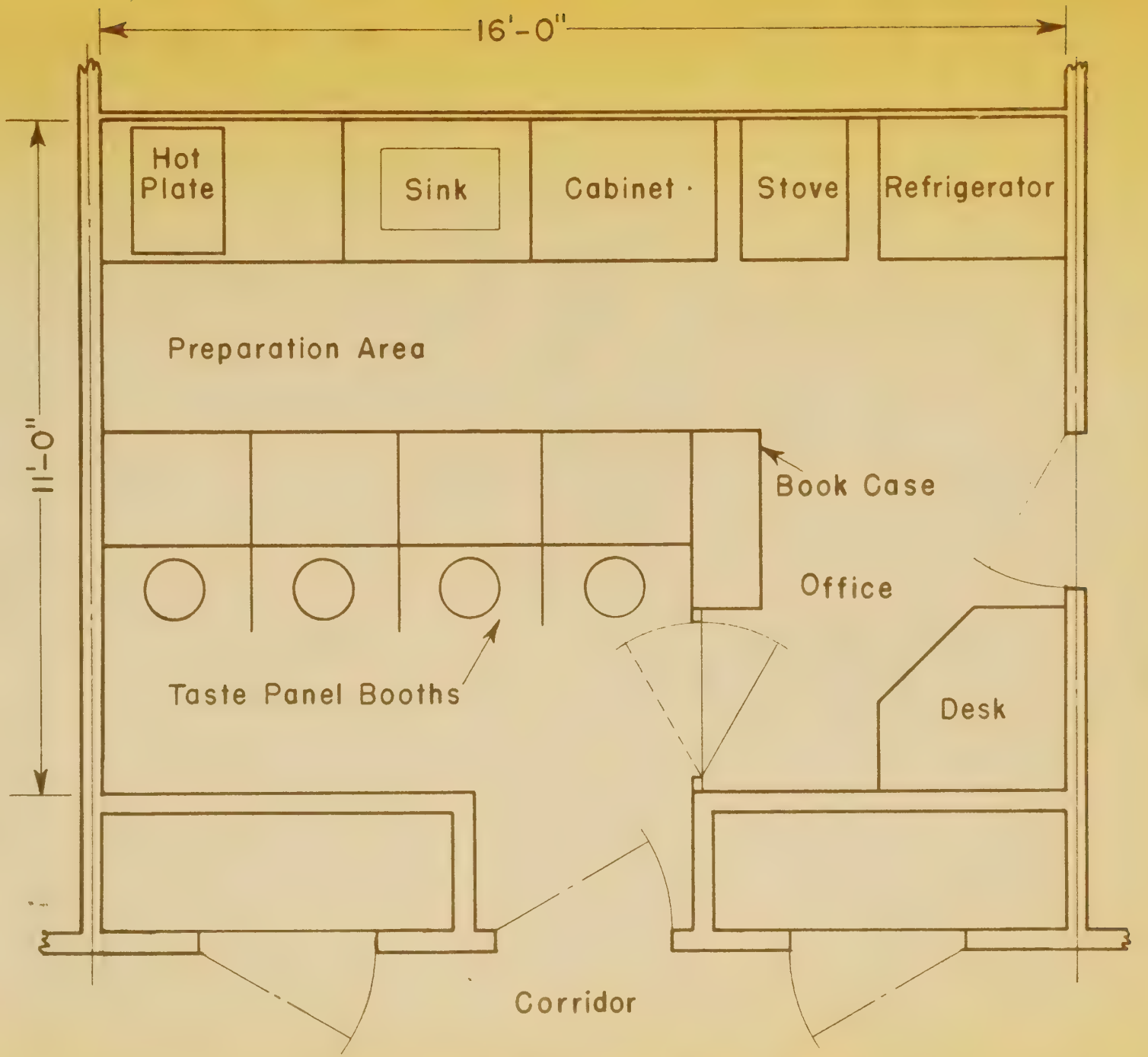
Literature Cited

1. Anonymous.
1943. Experimental procedure for conducting taste and smell tests. U. S. Dept. Agr. Bur. Human Nutr. Home Econ. Mimeographed release, November.

2. Bengtsson, Kjell.
1947. Taste testing as a method of analysis. IV. Selection of taste panel. Svensk Bryggeritid-skrift 62 (11): 311 (12): 333.

3. Gray, Philip P., Stone, Irwin, and Atkin, Lawrence.
1947. Systematic study of the influence of oxidation on beer flavor. Wallerstein Laboratories Commun. 10: 186.

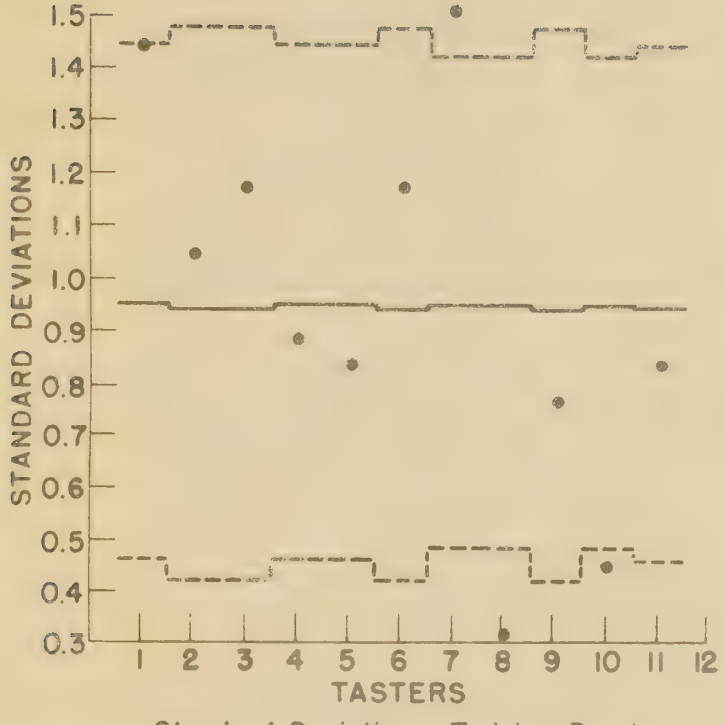
4. Roessler, E. B., Warren, J., and Guymon, J. F.
1948. Significance in triangular taste tests. Food Res. 13 (6): 503.



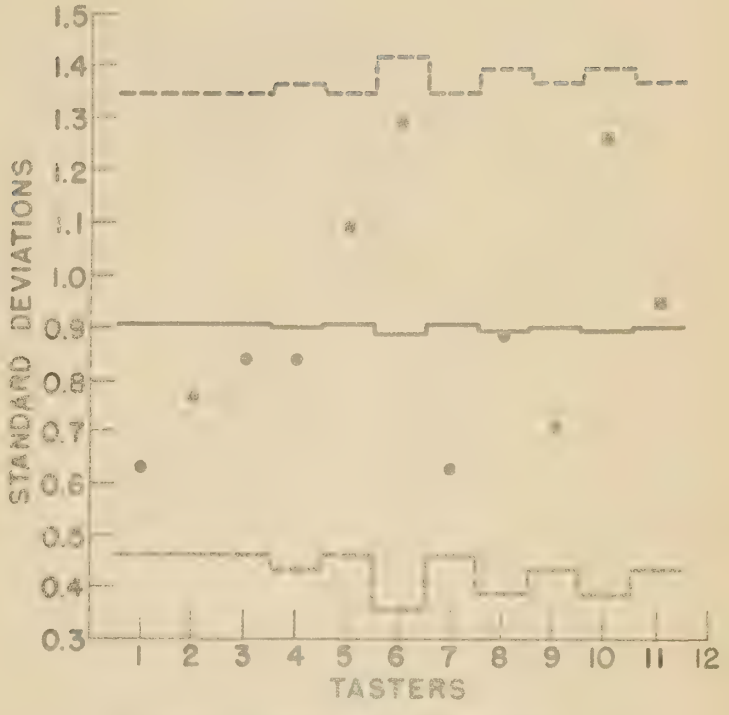
Floor Plan of Taste Panel Room

Fig. 1

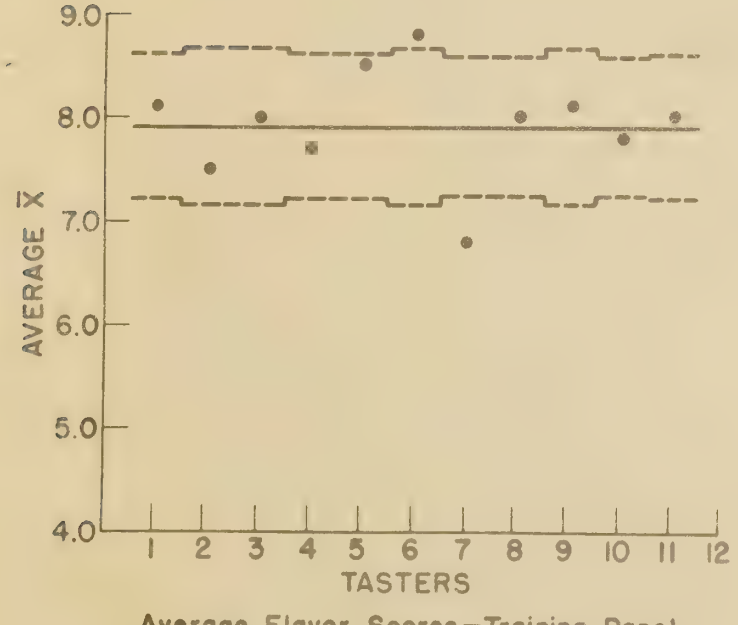
CONTROL CHARTS



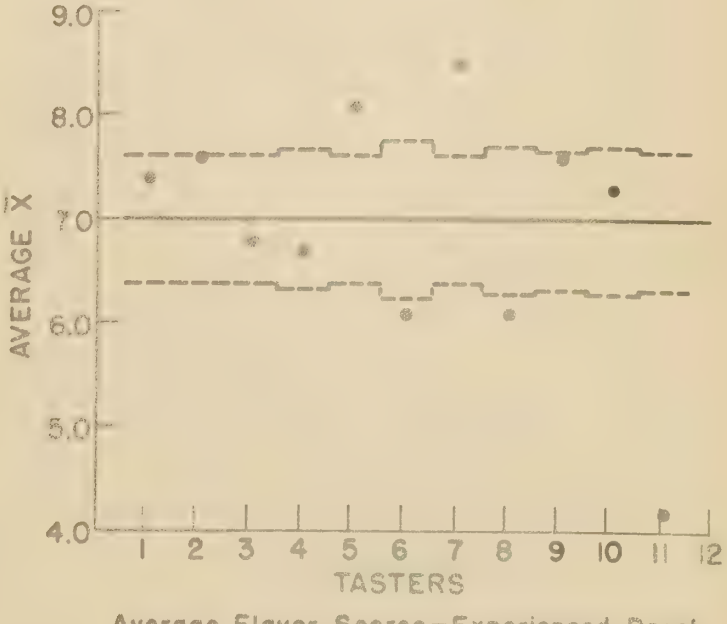
Standard Deviations—Training Panel



Standard Deviations—Experienced Panel



Average Flavor Scores—Training Panel



Average Flavor Scores—Experienced Panel

Fig. 2

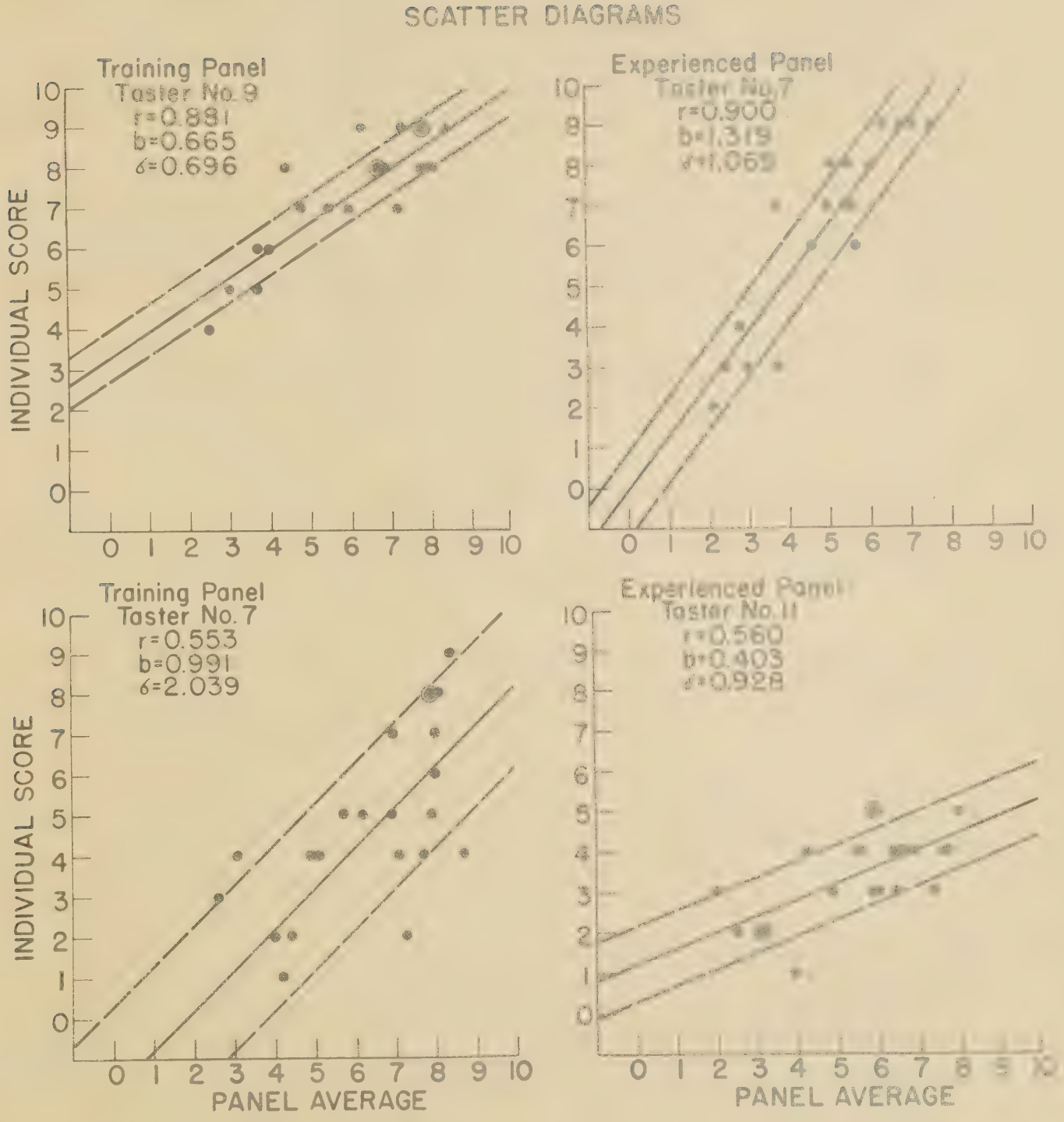

SCATTER DIAGRAMS
Training Panel
Taster No. 9
r=0.881
b=0.665
δ=0.696
Experienced Panel
Taster No.7
r=0.900
b=1.319
δ=1.069
Training Panel
Taster No. 7
r=0.553
b=0.991
δ=2.039
Experienced Panel
Taster No.11
r=0.560
b=0.403
δ=0.928
INDIVIDUAL SCORE
INDIVIDUAL SCORE
PANEL AVERAGE
PANEL AVERAGE

Fig. 3

IIIa. DESIGN OF EXPERIMENTS*/--Cyril D. Evans, NRRL

In many respects, the evaluation of edible oils presents fewer difficulties than the evaluation of many foods. Relatively few varieties of soybeans enter the trade, although there are many varieties known. In contrast, we are told that there are about 6,500 varieties of apples and some 500 varieties of potatoes. Varietal differences plus all those factors of cooking and/or sample preparation contribute to the difficulties of organoleptic evaluation. If you would think for a moment about the numerous problems involved in selecting and evaluating a new variety of cooking apple, you will see my point.

It is my own feeling that many of the differences found between edible soybean oils are not the result so much of varietal difference as they are the result of conditions brought about during processing. These conditions are tangible and are, to a greater or lesser degree, under our control. Thus we believe that the application of organoleptic evaluation can and will aid in the improvement of the quality of soybean oil.

The type of taste panel established depends, of course, upon the object of the investigation. For convenience, it is usual to divide the types of taste panels into four categories: (1) Consumer; (2) Control: (3) Analytical; (4) "Clinical" (miscellaneous, screening, trouble shooting, etc.).

The consumer type of taste panel deals with analysis of preference and encompasses all the problems and peculiarities, likes and dislikes of the human race. The problems of this type of panel are many and interesting, but not of main interest to our meeting today. It has been said that the "liquor trade" is the only large producer not interested in consumer panels.

The control panel is one designed for the maintenance of certain levels of quality within specifications. This type of operation is now widely used, although the introduction of the control chart technique is a rather recent development. Some of you are undoubtedly using this latter type of quality control.

Under such a control procedure, a set of specifications are made up. The product is then inspected to see if it measures up to the standards. With the objective type of specification, this procedure works fine, but with the subjective type, standards and specifications are more difficult to establish. Usually this latter type of evaluation is done by only one or two experts of long experience.

At NRRL we are most concerned with the analytical type panel. Judges are asked to find differences between two or more samples. Tasters may also be asked to describe discovered differences. As Dr. Dove stated, "It is the ideal to qualify and to standardize the measurement of the subjective responses so that they may then be combined with objective tests whenever possible." When differences are described, they are of greater value when the results are in a form suitable for statistical evaluation.

---

*/ Presented at the *Conference on Taste Panel Procedures and Methods of Evaluation for Soybean Oil and Soybean Oil Products* held at the Northern Regional Research Laboratory, Peoria, Illinois, on December 9, 1949.

As has been pointed out many times, the more sensitive tasters are the most valuable for a panel. The size of analytical panels may vary anywhere from 2 or 3 people, to 24 or 25. Most panels, however, will have perhaps about 10 or 12 tasters. For an analytical panel, tasters having a high coefficient of regression and a low standard of error are desired. For a panel of a small number of tasters, a high degree of agreement must be reached before significance in results is attained. From an over-all consideration, about 10 tasters constitute what is probably the most desirable size of panel for the paired sample technique. Later, tables will be presented showing the agreement that must be attained for panels of different sizes to show statistically that there are significant differences between samples.

The method of scoring is very important. In setting up an experiment, a good deal of time should be spent in devising the scoring system and developing a suitable score chart. Scoring may give considerably more information than ranking, or a simple "yes" or "no" answer. Ranking in order is one of the common scoring methods, but ranking alone is often misleading unless comparisons are made between samples that are of approximately the same value. If an unknown sample of good quality is compared against a standard of poor quality, one obtains a relatively high score for the unknown sample. If the comparison is made with an exceptionally high quality standard, then a disproportionately lower score is obtained for the unknown. To overcome some of these difficulties, the use of a series of standards has been employed to evaluate an unknown. Dr. Handschumaker has made excellent use of such a system in evaluating soybean oil shortenings. Others have used one standard in a series of unknowns. It is also common practice to use a duplicate pair in a series of rankings. Where such a system is applicable, differences will be shown and incorrect judging scores can easily be eliminated.

Details of the paired sample technique in which only two samples are evaluated at any one tasting are known to most of you. The main disadvantage is that it is time-consuming when a large number of samples are to be evaluated. It is our belief that a large number of oils cannot be tasted at one time with any high degree of accuracy. A single poor sample frequently dulls the taste senses to such an extent that further reliable grading is impossible. Another disadvantage is the large number of tasters required or, alternately, the large number of repeat tastings necessary to establish a significant difference between two samples. This is because of the high probability (0.5) of chance selection offered by paired sample technique. For three samples, each of two different oils, the probability of correct chance selection drops to 0.125.

In many tests using the paired sample technique the Northern Regional Research Laboratory panel will divide itself about equally in scoring A above B and B above A, and there will be very few members who will say that A is equal to B. Occasionally, a set of data is also found in which, although almost all members will score A above B, analysis of results will show no statistical difference. This is due to a high standard of deviation, (scores may vary from 2 to 8 in a single sample) and probably indicates samples of odd or unusual character.

Score sheets should never be highly complicated; an attempt should not be made to evaluate innumerable factors. Possibly two to four factors are most desirable. Scoring systems of 3, 5, 7, 10, 11 or more units are all being used. Many people like a so-called balanced scoring system with an even number of units above and below the satisfactory level.

At this point, I would like to give a short explanation of J. W. Hopkins' scoring system Dr Hopkins is the biometrician for the National Research Laboratory, Ottawa, Canada. He has developed the following scale which is used both for a trained taste panel and for the consumer response investigations:

			Attribute			
			Odor	Flavor	Color	Texture
	Gross	+5				
	Very decided	+4				
(Excess)	Decided	+3				
	Moderate	+2				
	Very slight	+1				
(Ideal)	None	0				
	Very slight	-1				
	Moderate	-2				
(Deficiency)	Decided	-3				
	Very decided	-4				
	Gross	-5				

Dr. Hopkins maintains that one is asking for the judges' or consumers' responses. Thus, in their responses, the ideal or zero point is established from the scale. For evaluation of soybean oil or cottonseed oil, I do not believe this scoring system has any great advantage over the system we now use. However, if we were to evaluate corn oil, and the standards of color, odor, and taste were to be maintained as they now are, one could use such a system to considerable advantage.

Another evaluation problem which we now have at the Laboratory is in the utilization of soybean flour. I have suggested a scoring system similar to Dr. Hopkins' in the evaluation of soy flour bakery products; it seems preferable to the one we are using for oil work.

Score sheets should be very specific and, in my opinion, comments should always be allowed and even asked for. Before using a new scoring sheet or system, panels should be well informed as to just what is wanted or desired in a particular series of tests.

In any discussion on taste panels, a few words should be said about maintaining the morale of the panel. The maintenance of a high spirit of cooperation and interest is the most difficult and also the most important job in the entire operation of the taste panel. The first qualification of a good taster is interest in doing the job.

Taste panel members cannot be regarded as guinea pigs; they are artisans in their own right. They should be fully informed of the objectives of the experiment and the results when the experiment is completed. If tasters are interested and willing to cooperate and devote their time to tasting, those in charge must seek to maintain their interest in every possible way An uninterested taster is sure to become a liability. Time spent in training and developing a taster is wasted if the individual loses interest in the evaluation work.

III b. SIGNIFICANCE OF TASTE TESTS FOR SMALL ORGANOLEPTIC PANELS*/

C. D. Evans and Earle Lancaster, NRRL

Statisticians have accepted results with probabilities of less than one chance in twenty as being significant. This is the so-called 5 percent level of probability and results within this limit are customarily designated by a single asterisk (*). Results that do not reach this level of probability are spoken of as non-significant and are designated by a single plus mark (+). If the probability by chance is so low as to be less than 1 in a 100, this is called the 1 percent level. Such results are considered as being highly significant and are designated by two asterisks (**) Three asterisks designate probabilities of chance of less than 1 in a 1000.

Under the conditions encountered in ordinary taste testing, probabilities can be rather easily calculated for these experiments. It is the significance of the results, however, that are of interest and of value in the operation of a taste panel.

What do we mean by significance? In taste testing of 2 samples it is obvious that we are required to know how many tasters must agree that sample A is different than sample B before we can have any confidence in the answer. If there is an actual difference between samples A and B, then what is the probability that our taste panel could give the same results on the basis of pure chance? Assuming any probability of less than 1 chance out of 20 (5 percent level) as being accurate for our purpose, we can then calculate how many of the tasters must agree in order to establish that a difference exists between any two samples. If a significant difference cannot be demonstrated we are still not permitted to state that the two samples are identical. It is more accurate to state that no difference has been demonstrated or established by our test.

There are, of course, several ways in which one can present two samples to a taste panel to find out if a difference exists. The simplest method would be to take only one sample of A and one of B and then have the tasters *rank* these two samples, that is, decide which sample is better, without permitting a "no difference" response. If there is no apparent difference, then A, on the average, would be ranked ahead of B by one-half of the tasters. This test has the same probability (1/2) as in the tossing of a coin, i.e., a 50-50 chance.

In *scoring* oils, however, one may come up with three possible answers. $A > B$ or $A = B$ or $B > A$. This is one difference between ranking and scoring. In scoring, any significant difference between sample scores is calculated by the usual statistical method of variance analysis. Thus, the tasters must be forced to select A or B as the better sample if the (50-50) probability tables are used to establish results. This holds true as well in the triangular test (where an answer of "no difference" also is not allowed) if the probabilities of 1/3 one used to establish a significant difference.

The triangle test offers a second method to distinguish differences between samples A and B. Here two samples of A are presented and one of B (or two of B and one of A), and the tasters are informed that a triangle test is being given and that two samples are identical.

---

*/ Presented at the ***Conference on Taste Panel Procedures and Methods of Evaluation for Soybean Oil and Soybean Oil Products*** held at the Northern Regional Research Laboratory, Peoria, Illinois, on December 9, 1949.

The judges are asked to select the two samples which are identical. By chance any person without tasting has the probability of 1/3 of picking out the correct pair. Samples could be sorted as A_1A_2-B (the correct answer), A_1B-A_2 or A_2B-A_1.

A third method to ascertain whether any differences exists between A and B is to give the taste panel two samples of A and two samples of B and then ask them to correctly pick out and identify each pair. Types of oils used should be fairly familiar to the tasters. The panel members must be told that there are two of each sample so that they will avoid reporting three of one type and one of the other, or all of one type. In selecting the two correct pairs in this case a random chance selection of samples will give a probability of only 1/6. The possible chance selections are as follows:

A_1A_2 and B_1B_2 *	A_2B_1 and A_1B_2
A_1B_1 and A_2B_2	A_2B_2 and A_1B_1
A_1B_2 and A_2B_1	B_1B_2 and A_1A_2

* The correct selection.

If, however, one asks the taste panel members only to pick out the two pairs of samples, and not attempt to identify them as A's or B's, then the probability is 1/3. Both AA - BB and BB - AA are now correct answers.

Similarly, three samples of each type of oil could be given the judges to match the triplicate samples and identify the two sets. By increasing the number of identical samples in the matching and identification technique from a single pair to two pairs of different oils, the probability has been dropped from 1/2 to 1/6. Going to three samples each of two different oils, the probability of chance selection drops to 1/20 (1). The presentation in the latter case of such a large number of oils (six samples) for tasting would be impossible. However, they could be ranked by smell alone.

Another type of test which we have employed for the identification of simulated oils is as follows. The judges are presented with three samples, and they are asked to identify them as either cottonseed or soybean oil. They are told that they may be either cottonseed oil or soybean oil or all one or the other. Here again three samples are presented but the probability of chance selection for the correct answer is now much less. It is 1/8 since the selection of each sample is 1/2 (only two choices possible for each sample). As there are three samples, the probability is easily calculated (1/2 x 1/2 x 1/2 = 1/8). In this presentation, the identification of the first sample has no effect on the selection of the other two remaining samples and the probabilities are derived by simple multiplication. The possible combinations which can be selected are:

AAA	AAB	ABA	ABB
BAA	BAB	BBA	BBB

Note that this technique is different than in the triangle test, where choice of second and third samples is not independent of the first choice. In the triangle test, the probability calculation is based on the selection of combinations of three things taken two at a time.

$$C = \frac{N!}{P!\,(N-P)!} = \frac{3!}{2!\,(3-2)!} = 3$$

A similar type of experiment is represented where two oils are presented to the panel to be identified as soybean or cottonseed oil. In this case, the probability is only 1/4 (1/2 x 1/2 = 1/4); the possible selections the taster can make being AA, AB, BA, or BB. Here a random selection can be correct on the average of once out of every four trials.

The probability of chance selection for a single taster has been developed and it is a simple function of permutations and combinations. From the foregoing examples, it can be seen that probabilities varying from 1/2, 1/3, 1/4, 1/6, 1/8, 1/20 or higher can be selected when evaluating two samples depending on the manner in which the samples are presented to the taste panel.

When more than one taster is involved the same considerations of probability apply but the larger number of tasters must be taken into account. Expansion of the binomial expression makes it possible to calculate the number of possible compound probabilities under these conditions.

If p = probabilities of success

q = probabilities of failure

n = number of independent events

$$p + q = 1$$

$(p + q)^n$ = compound probabilities

Consider the possibility of a probability of 1/2 for each of three tasters. Then the expansion of the expression $(1/2 + 1/2)^n$ will give the compound probabilities for three independent events.

$(1/2 + 1/2)^3$ expands to $1/8 + 3/8 + 3/8 + 1/8$

Thus, we have 1/8 of a chance of all three tasters making a correct selection, and 3 chances out of 8 of making 1 error, also 3 chances out of 8 of making exactly 2 errors and, of course, 1 chance out of 8 of making three errors, that is, of every taster selecting the wrong sample.

Now for ten tasters the expression for a probability of 1/2 becomes $(1/2 + 1/2)^{10}$ and this on expansion gives

$$\frac{1}{1024} + \frac{10}{1024} + \frac{45}{1024} + \frac{120}{1024} + \frac{252}{1024} + \frac{210}{1024} + \frac{120}{1024} + \frac{45}{1024} + \frac{10}{1024} + \frac{1}{1024}$$

This indicates that for a panel of ten tasters, the probability of all ten selecting the correct sample by chance is 1 in 1024. We also have 10 chances out of 1024 of making exactly one wrong selection, and 45 times out of 1024 of making exactly two wrong

selections Thus, the probability of making either none, one, or two errors is the sum of these probabilities, that is:

$$P = \frac{1 + 10 + 45}{1024} = \frac{56}{1024} = .0537$$

The probability of a panel of ten members not making more than two errors by chance is only slightly higher than required for a significant result at the 5 percent level (05) In other words, to allow a maximum of two wrong answers in the selection of sample A over sample B the panel must have a minimum of eleven members to indicate a significant difference between them.

Table I (see page 5) has been constructed for the probability of 1/2 per taster for one to fifteen tasters. Values of the binomial coefficient for expansions of 1 to 20 is tabulated in most technical handbooks.

From data in Table I the number of errors allowed for any number of tasters can be calculated by summing up the appropriate figures, as was done in the example for ten tasters.

In Table II (see page 6) are tabulated the number of errors that may be allowed for a given number of tasters and a given set of probabilities These values have been calculated for a significance above the 5 percent level.

The number of errors allowed for significance in the case of probabilities of 1/3, the binomial expression $(1/3 + 2/3)^n$ is expanded and the appropriate calculation made Similarly, for a lower probability such as 1/8 the chances of guessing correctly are given by the binomial $(1/8 + 7/8)^n$ which gives the compound probabilities expected for n number of tasters. Roessler (2) has published a table for significance in triangular tests for seven to 100 tasters. Gary, et al., (3) has also plotted similar results for a large number of samplings.

Table I

BINOMIAL EXPANSION FOR THE PROBABILITY OF 1/2 FOR 1 TO 15 TASTERS

Exact No of Errors:	Number of Tasters														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2		1	3	6	10	15	21	28	36	45	55	66	78	91	105
3			1	4	10	20	35	56	84	120	165	220	286	364	455
4				1	5	15	35	70	126	210	330	495	715	1,001	1,365
5					1	6	21	56	126	252	462	792	1,287	2,002	3,003
6						1	7	28	84	210	462	924	1,716	3,003	5,005
7							1	8	36	120	330	792	1,716	3,432	6,435
8								1	9	45	165	495	1,287	3,003	6,435
9									1	10	55	220	715	2,002	5,005
10										1	11	66	286	1,001	3,003
11											1	12	78	364	1,365
12												1	13	91	455
13													1	14	105
14														1	15
15															1
2^n	2	4	8	16	32	64	128	256	512	1,024	2,048	4,096	8,192	16,384	32,768
5% of 2^n	0.10	0.2	0.4	0.8	1.6	3.2	6.4	12.8	25.6	51.2	102.4	204.8	409.6	819.2	1638.4
1% of 2^n	0.02	0.04	0.08	0.16	0.32	0.64	1.28	2.56	5.12	10.2	20.5	40.9	81.9	163.8	327.7

Table II

Number of errors permitted by a taste panel to establish a significant difference at the 5 percent level, when the probabilities of random selection are 1/2, 1/3, 1/4, 1/6, and 1/8.

Number of tasters on panel	Probabilities of 1/2	1/3	1/4	1/6	1/8
	Errors allowed cannot exceed				
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	1
4	0	0	0	1	1
5	0	1	1	2	2
6	0	1	~~1~~2	2	3
7	0	2	2	3	4
8	1	2	3	4	~~5~~4
9	1	3	4	5	~~6~~5
10	1	3	4	5	6
11	2	4	5	6	7
12	2	4	5	7	7
13	3	5	6	7	8
14	3	5	7	8	~~10~~9
15	3	6	7	9	~~11~~10

REFERENCES

1. Wood, Eric C. Jour. Soc. Chem. Ind. *68*, 128 (1949).

2. Roessler, E. B.; Warren, J.; Guymon, J. F. Food Res. *13*, 503 (1948).

3. Gray, Phillip; Stone, Irwin; Atkin, Lawrence. Wallerstein Comm *10*, 183 (1947).

IV METHODS OF EVALUATING SALAD OILS AND MAYONNAISE*/ --

Dr. S. Jack Rini, Kraft Foods Company

(Written from notes taken during the Conference)

In grading salad oils, evaluation problems are encountered which are somewhat different in character than those which have been discussed for unblended oils. Corn, soybean and cottonseed oils are combined in varying amounts to produce salad oils. Individual bias is one complication. Certain tasters have been found to have either "corn oil phobia," or a distinct dislike for some other oil in the series mentioned. Thus, in any random cross section of taste panel membership, there will almost always be some bias for or against a particular constituent. For the sake of more reliable evaluation it is necessary to weed out biased members, regardless of whether that bias is positive or negative with respect to a particular constituent.

In the evaluation procedure being employed by Kraft, considerable heating of the sample gives a good means of detecting odor differences which cannot be discerned under any usual conditions In this procedure, samples are heated to 180° 190° F , and ratings of the samples are made by both taste and smell.

Kraft has no definite taste panel arrangement as yet. A panel is being organized, however, and persons are being screened for membership on the panel. Problems which will be encountered are expected to be as varied as the products which will require the services of a taste panel. Several different panels may be required to cover Kraft's needs, each trained in one of the several different fields required.

Evaluation of mayonnaise also presents special problems. Both odor and flavor are checked. It has been learned, however, that presence of acids, spices, and other flavoring factors in mayonnaise affect considerably the acuteness of flavor response. Likewise, starch in salad dressings plays an important part in the flavor response As could be expected, changes in relative amounts of these various ingredients affect the flavor response obtained when attempts were made to evaluate oils in such products. The consumer preference angle comes in here. In samples of mayonnaise incubated at 100° F., a peculiar eggy flavor shows up during the usual weekly or monthly gradings. There is a tendency to blame the oil ingredient for any of these shortcomings.

To hold flavors of certain products as nearly constant as possible over a period of time, the samples are stored at 26° F. It is impossible to use lower temperatures because mayonnaise or salad dressing cannot be allowed to freeze. This rules out the possible use of a deep freezer suggested by another speaker.

Certain direct and practical methods of testing vegetable oils being used as ingredients of food products have been tried. The oil being tested has been used for popping corn, frying french fries, or making doughnuts. Differences are observed in the food products produced. Any undesirable characteristics of vegetable oils usually show up under these conditions.

Mayonnaise samples are tasted in two operations. Using a paper spoon, a portion from the top of the sample and another from the interior are tasted separately. The flavor of the top portion indicates the flavor development to be expected from oxidative effects and the cap liner. Samples are ranked in accordance with the results of these tests.

---

*/ Presented at the *Conference on Taste Panel Procedures and Methods of Evaluation for Soybean Oil and Soybean Oil Products* held at the Northern Regional Research Laboratory, Peoria, Illinois, on December 9, 1949.

V. METHODS OF EVALUATING SHORTENING*/

P & G Method of Evaluating the Flavor Stability of Soybean Oil Products—R T Clause, Procter and Gamble Company:

The Procter and Gamble method of evaluating the flavor stability of soybean oil is that described by Sanders in 1944 (J. H. Sanders Predicting the Flavor Stability of Soybean Oil *Oil and Soap* December 1944). He pointed out that the flavor quality of normally refined soybean oil is related to its bleach color about as follows:

Flavor Quality	Approximate Lovibond Red Bleach Color	
	6% English Earth	4% Activated Earth
High Quality	3.0 max.	1.5 max.
Intermediate Quality	3.1 to 5.5	1.6 to 3.5
Low Quality	5.6 to 8.5	3.6 to 6.5
Inedible	Above 8.5	Above 6.5

Use of a grading scale such as the above will not result in perfect segregation of soybean oil receipts into quality classes but two thirds of the oils can be accurately classified. The other one-third will fall close to the arbitrary dividing line between the grades. For our edible products, we choose oils according to the grade into which they fall according to the bleach color, and this gives us a fairly accurate picture of the flavor stability of such oils.

For actual evaluation of the flavor stability the edibility test is used. The oil to be checked is alkali refined, settled overnight then decanted from the foots and clarified by filtration. The clarified oil is bleached with a commercial bleaching earth. A portion of the bleached oil is set aside for deodorization, and the remainder is hydrogenated to 75 I.V. - about the range for shortening and margarine The refined and bleached oil and the refined bleached and hardened oil are next deodorized on a one pound scale The deodorized oil is filtered into a new tin can It is flavored as freshly deodorized oil, then stored at 140°F. The unhardened oil is again flavored after 48 hours, and the hardened after 72 hours. The flavor quality of the aged samples is a measure of the flavor stability of the oil

The same procedure as is used to check the flavor stability of edibility test samples is used to check the flavor stability of our finished products, i.e deodorized products are flavored fresh and after aging at 140°F Normally one person qualified by years of experience makes the flavor ratings Occasionally a panel of several people will grade samples with him as a check of his ratings.

In rating any product from a flavor standpoint a graded scale is employed For this discussion, let us consider a grading scale in which 10 is the best quality and 1 is the poorest To have a 10 flavor grade, a sample must be almost bland. It should have no more than a trace of a type flavor which is not objectionable Normally, freshly deodorized oil processed under optimum conditions from good quality crude will fall in the 9 to 10 flavor grade range, and will age at 140°F. to a flavor in the 7 to 8 range.

---

*/ Presented at the ***Conference on Taste Panel Procedures and Methods of Evaluation for Soybean Oil and Soybean Oil Products*** held at the Northern Regional Research Laboratory, Peoria, Illinois, on December 9, 1949.

Very few soybean oils will fail to show reversion after the 140°F. aging period, however. It is not the fact that a reversion flavor develops that is objectionable, but the type of reversion flavor which develops.

Following is a table showing the type flavor and the flavor grade that can be expected in the hydrogenated oil from the various quality crudes:

Flavor Quality	Aged Flavor Grade (3 days at 140° F.)	Type Flavor
High Quality	8 min.	Sweet to slight buttery.
Intermediate Quality	6 - 7	Buttery to fishy.
Low Quality	4 - 5	Pronouncedly fishy.
Inedible	Under 4	Very fishy, oxidized, painty.

The Evaluation of "Off-Odor" in Hydrogenated Soybean Oil Shortening—*Edward Handschumaker, Lever Brothers Company*:

The problem of estimating the extent of "reversion", or more properly "off-odor", development in hydrogenated soybean-oil-containing shortenings resolves itself into two separate measurement problems. The first comprises the evaluation of the consumer's reaction to this aroma, and the second involves the routine laboratory tests required to control the quality of production and to appraise the effect of research efforts.

With respect to the consuming public, a number of interesting questions could be raised, among which are: How many people object to this "off-odor"? What is the threshold intensity at which the majority of consumers will notice the aroma? Is this odor more noticeable in some products than in others? Can the observed preferences be classified geographically?

A scientific approach to this problem would embody the use of the most up-to-date population sampling techniques. A number of difficult but not insurmountable problems would certainly arise. Suitable lots of hardened soybean oil and cottonseed oil would have to be prepared and blended. Fried products would have to be prepared according to specifications in several places. Local product preferences should be considered. All told, quite an organization would be required to conduct the surveys. Perhaps most of the individual tests comprising the surveys would need to be replicated at least once. The results of such a study would certainly be very helpful to the industry as a whole, and it would undoubtedly require the pooled resources of the whole industry or of the government to pay for the program. The basic information provided by the comprehensive consumer study would also be very useful as a basis for the calibration of smaller local panel groups, which could then be considered to represent the consuming public.

However, if we are willing to assume, as some have, that the "off-odor" is generally objectionable to the majority of consumers and that a small, randomly selected panel of our own employees or neighbors can safely be considered to represent the population of the country as a whole, the problem can be greatly simplified. Under these conditions two panels could readily be established to carry out the work.

The first and comparatively large group of 100 or more people would provide the consumer opinion estimates. Fried products, like doughnuts and potato chips, would be submitted to them in pairs for a declaration of preference. As nearly as possible these pairs ought to differ from one another only with respect to the shortening used in their preparation. One sample of each pair should represent a control fat considered to be of acceptable quality. Neither the control nor the test sample ought to be identified in any way. Each test should probably be replicated at least once to permit correcting the estimates for individual decisions arrived at by guessing. Replication would also provide an estimate of the experimental error associated with the test.

The panel evaluation should be conducted in such a manner that each individual preference expressed was not influenced by suggestions from others. The questions asked ought to be few in number and carefully worded to avoid bias. Each individual participant might reasonably be asked if he regularly consumed the product being tested, whereupon he would be permitted to register his opinion.

He might prefer one or the other of the pair or have no choice between them. If our original assumptions regarding the representativeness of the panel are reasonably close to the truth, the foregoing test program would serve to keep us fairly well informed about the consumer's opinion.

The problem of measuring small odor differences accurately and with the precision required for routine production control or the evaluation of research efforts is quite independent of the particular technique which is used to obtain the consumer's preference. In this case it is desirable to eliminate all possible confounding factors and confine our attention entirely to the shortening itself.

A small panel composed of the most sensitive observers which can be found will be most effective in this phase of the program. The shortening must be exposed to standardized elevated temperature conditions to accelerate in a reproducible manner the development of the objectionable odor. A suitable scale of controls should be provided against which a test sample can be graded. The observations must be made by smelling, because olfactory fatigue will usually set in before the contemplated test can be completed if tasting is practiced. Restricting the evaluation to smelling introduces no real irregularity, however, because the objectionable quality factor associated with heat-exposed soybean oil shortening is quite definitely an aromatic agent.

A detailed method embodying the ideas expressed will be found in my published paper, *A Technique for Testing the Reversion Properties of Hydrogenated Soybean Oil Shortenings*. This paper appeared in the Journal of the American Oil Chemists' Society, 25: 54-56 (1948). The reported test can probably be simplified in many instances without sacrificing very much information. I believe the panel could be reduced to a single individual provided he possessed a sharp olfactory sense and if the rest of the restrictions imposed by the method are strictly observed. In this case converting the ratings to scores and the subsequent statistical analysis would, of course, be quite unnecessary.

I believe the comprehensive consumer test described in the first section of this paper is the most important of the three techniques mentioned. Without the data supplied by this test, the significance of the results obtained with the other two panel techniques

is difficult to appraise quantitatively. However, once a clear picture of the consumer's opinion has been recorded, it would be a comparatively simple problem to set up reliable small local panels to represent the whole population. The task of establishing the small panel of expert observers required for the laboratory evaluations is probably the easiest of all. In this instance it is important to avoid personal bias and to closely reproduce the test conditions.

For a good practical exposition of the fundamental principles involved in flavor panel work, I recommend Washington Platt's excellent paper, *Some Fundamental Assumptions Pertaining to the Judgment of Food Flavors*, which appeared in Food Industries, 2: 237-249 (1937).

Procedures and Methods Used by Wilson & Company in Organoleptic Evaluations of Shortenings and Their Flavor Stability——H. T. Spannuth, Wilson and Company, Inc.:

Our procedures can best be described in general terms as using small panels of "experts" with multiple checking. Expediency in getting useful opinions for arriving at what we expect to be correct decisions tends to predominate over the scientifically controlled tests of "statistically proven conclusions". The latter is reserved for major changes in formulas and processes. In general, the plan is for the Home Economics Department to conduct such tests. Being aware of the subjective nature of organoleptic testing, we have strived to secure abusive tests through which one may get "leverage" to confirm opinions. The work is separated into control and research.

A. Control

1. Raw materials.
   a. Plant check. b. Laboratory check.
2. Materials in progress.
   a. Plant check. b. Laboratory check.
3. Finished products.
   a. Plant check. b. Laboratory check. c. Shortening tasting panel (with or without aids of substantiating test).

B. Research

1. Difference tests. 2. Acceptance tests. 3. Performance tests.
4. Shortening taste panel A(3) (c). 5. Special taste panel of Home Economics Dept.

The plant foreman is responsible for product produced even though the Research and Technical Division carries a joint responsibility which is more of a nature of judicial review, thus a check and balance system exists. Problem cases are carried to either the regular tasting panel or, where desirable, to an enlarged panel as will be explained later.

Normalcy test: The plant and laboratory personnel are alerted for any tell-tale signs of product being different as revealed by taste, odor, appearance or unusual chemical analysis such as clarity, FFA, peroxide value, AOM etc. Observations during the running

of the AOM tests are considered useful. Samples under suspicion are referred to the Central Control Laboratory where examination is made by an expert. He calls on the panel if the case is not clear cut.

In conducting the above tests, the worker and/or the analyst checks flavor and odor, and reports abnormalcies immediately to the supervisor. It is common practice for the person in charge to select those whose opinion he considers most valuable to rate product in question in a blind test. Difference tests are frequently used which will be explained later.

Shortening Tasting Panel: A tasting panel exists for each type of product. The one to be described here is for shortening. The test is run on a 3-pound sample filled and sealed with the regular batch number and is inspected after complete testing program and analysis are available. The panel normally consists of four expert tasters - one each from the Central Control Laboratory, General Superintendent's Office, Research Bakery and the Research Department. Alternates are available for substitution or inclusions so that the panel may include as many as eight but usually four. Since only bland shortenings are tasted it is possible for such a panel to inspect a relatively large number at one time, such as 30 to 50.

The practice of having each person record his individual secret rating was discarded in favor of each taster making known his rating. If there is lack of agreement, sample is rechecked. In case no agreement is then reached, difference is so recorded or a person is called in to split a tie. An attempt is made to get a one-value rating and to pick out exceptional samples. The reason for favoring the latter procedure is that there is much less tendency for a person to consistently grade the shortening as average, which on our scale is fair. The scale employed is: Good - 10, Fair to Good - 9, Fair - 8, Low-Fair 7, Fair to Poor - 6, Poor - 5.

It was soon learned under the first system that there is considerable tendency for a person to rate all shortenings fair since this resulted in the least deviation from the average. Under the second system it was found that one was more prone to rate a shortening low-fair or fair-to-poor or fair-to-good since there were three persons checking and the sample could be rechecked by all with a final average rating before going to the next sample.

It should be fully recognized that good bland shortenings are being tested and that the object is fundamentally to pick out less desirable product with the view that corrective measures will be taken. Note that persons checking samples are in position to use information directly. All samples have previously been screened during processing and prior to packaging by the plant foreman and chemist and also after packaging.

Substantiating tests: The FFA may be used under a given set of conditions to indicate the effectiveness of deodorization. A well deodorized shortening will have a negative Kreis test. The peroxide value of a fresh shortening should be negative or only a trace. The active oxygen test on all hydrogenated shortening should be over 60 hours in the case of emulsified and general purpose types and over 80 for high stability types. The 140°F. incubator test is found useful supplement to the AOM test. The object is to compare length of time shortening can be held before development of objectionable flavor and odors. Samples are checked at two time intervals. Either 5, 10 or 15 day-intervals are normally used, depending on the shortening.

Performance Tests: The object of the performance test is to utilize a shortening in a product or products for which it is intended. The ones found most useful are potato chip, cracker, pie dough and prepared mixes (cake and pie). The potato chip is particularly good in view of the fact it indicates a factor of heat damage and permits the chip to be stored in light for observing characteristic light reversion. It is also well adapted to checking by accelerated incubator test at 140°F. as well as checking at room temperature, or better still at 92°F. The room storage tests are made in cellophane heat-sealed bags and/or in jars with loose cover caps. Crackers and pie crust can be tested similarly.

Tests are run using the doughnut machine made by the American Doughnut Corporation. Equilibrium potato frying tests are run using a Friolator. Care is exercised in maintaining a chip-to-oil ratio common to commercial machines.

Shortening Storage Test: Consumer-packaged shortening is stored both at 70°and 92°F. with an examination between the 6th and 12th month for the 70° samples and between the 3rd and 6th months for the 92° samples. Longer periods are used if desired.

Special Panels: Our Home Economics Department is charged with the responsibility of conducting objective and more elaborate tests, such as difference and acceptance tests on all products, including derived products. Approximately 65 individuals have been given acuity and threshold tests for various substances. Various procedures have been tried. Further proposed changes are to be evaluated. The work is in general instigated by other departments where specific questions for answering arise.